

LIGHTWEIGHT SUBJECT INDEXING FOR E-MAIL COLLECTIONS

Abstract of the Disclosure

A light weight subject indexing system including a candidate headword identification system for identifying candidate words in the subject line of a document which are not listed in a user modified common word list, a lexical context system for creating lexical context for an identified candidate headword, a ranking system for ranking all the candidate headwords identified for the subject lines of a document or message collection, and selecting among the ranked headwords for inclusion in an index based on that ranking, and an index creation system for listing candidate headwords selected by the ranking system.

0022423950